NEURAL INFORMATION PROCESSING SYSTEMS

LCDB 1.1: A Database Illustrating Learning Curves Are More Ill-Behaved Than Previously Thought

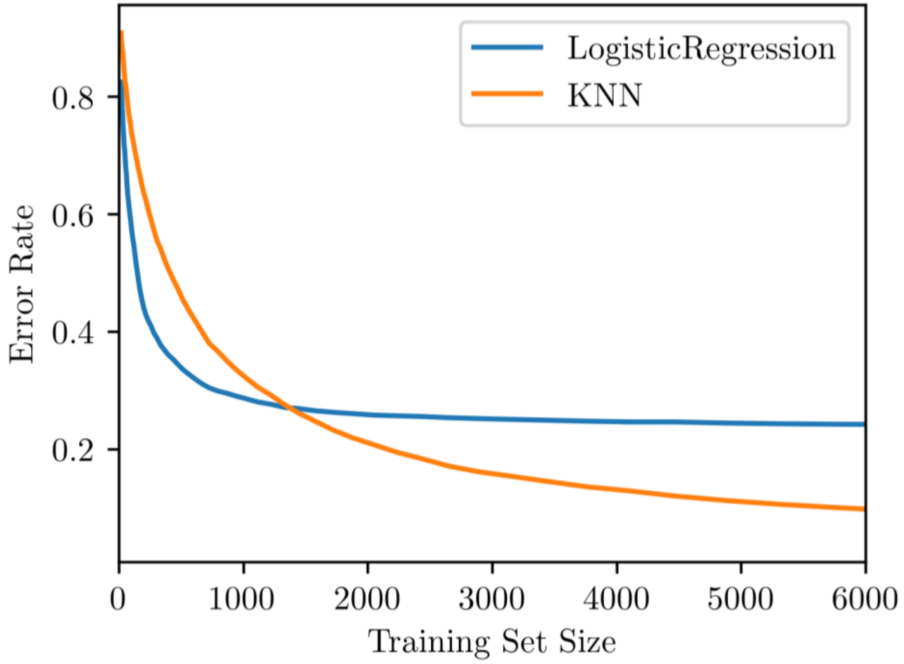
Cheng Yan[1], Felix Mohr[2], Tom Viering[1]

[1]Delft University of Technology, [2]Universidad de La Sabana

TU Delft

Motivation 1: Sample-Wise Learning Curve

- Richer Evaluation
- Model Selection
- Data Collection
- Performance Prediction



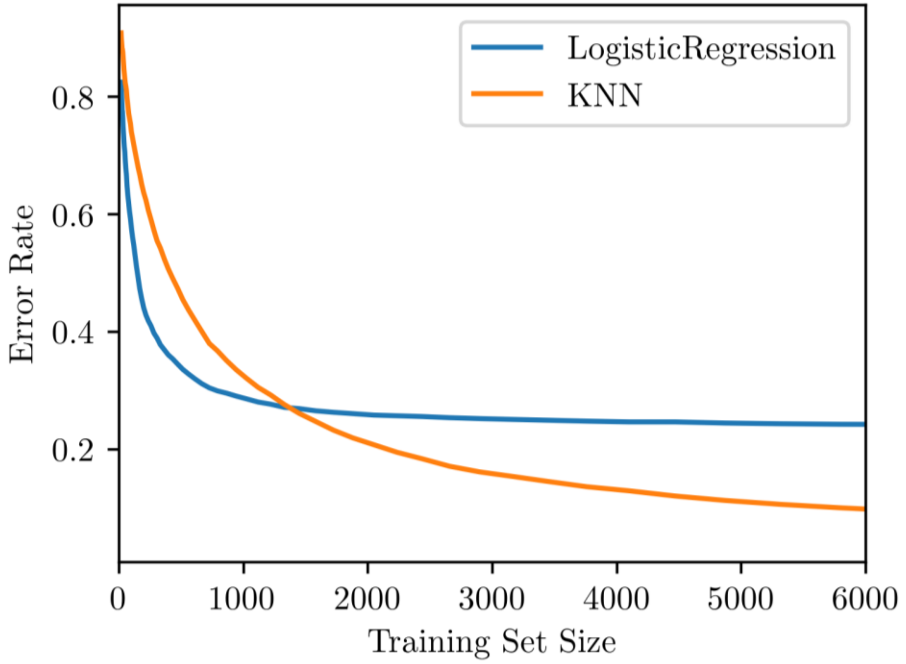
Motivation 2: The Shape of Learning Curves

- Monotone:

More data, better performance

- Convex:

Diminishing returns

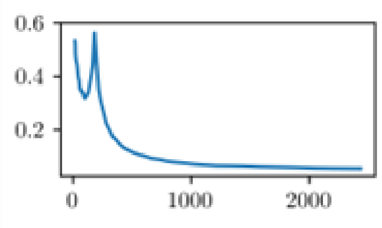
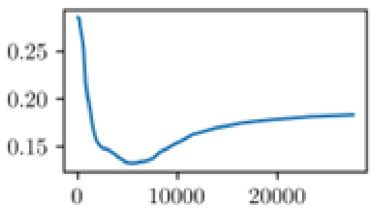
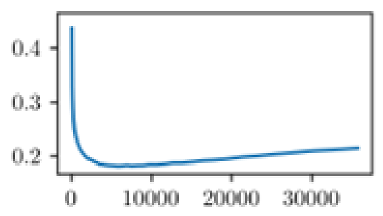
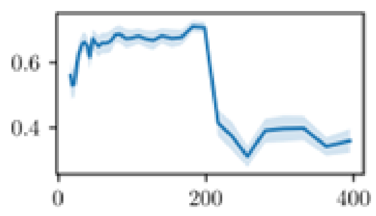
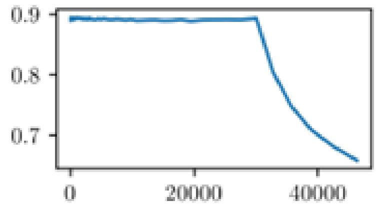


LCDB 1.1: Learning Curves Database 1.1

The largest database of sample-wise learning curves:

- **32 learners**: classical, ensemble, boosting, deep learning, foundation models
- **265 real-world** tabular classification **datasets** from OpenML
- **3 feature scaling** methods as pre-processing

The Shapes of Learning Curves Could Be Diverse

<table>
<tr><th>Peaking</th><th colspan="2">Dipping</th><th colspan="2">Phase Transition</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr><td>non-monotone
non-convex</td><td>non-monotone
non-convex</td><td>non-monotone
convex</td><td>non-monotone
non-convex</td><td>monotone
non-convex</td></tr>
</table>

Ill-Behaviors Vary Significantly Among Learners

- In total, around 15% of the learning curves are ill-behaved.

Learners (Abbreviation)	Ill-behaved	Learners (Abbreviation)	Ill-behaved
CatBoost	1.5%	Complement Naive Bayes (ComplementNB)	8.3%
Decision Tree (DT)	1.5%	Passive Aggressive (PA)	9.4%
TabPFN v2	1.5%	Mix Complement Naive Bayes (MixComplementNB)	10.2%
Extra Tree (ET)	1.9%	Mix Multinomial Naive Bayes (MixMultinomialNB)	10.6%
ensemble Gradient Boosting (ens. GB)	1.9%	RBF Support Vector Machine (SVM_RBF)	15.8%
ensemble Random Forest (ens. RF)	3.0%	Ridge Regression Classifier (Ridge)	17.0%
Stochastic Gradient Descent Classifier (SGD)	3.4%	Mix Gaussian Naive Bayes (MixGaussianNB)	21.5%
ensemble Extra Trees (ens. ET)	3.4%	Gaussian Naive Bayes (GaussianNB)	24.9%
Perceptron	3.8%	Multilayer Perceptron (MLP)	27.9%
K-Nearest Neighbors (KNN)	3.8%	Bernoulli Naive Bayes (BernoulliNB)	28.3%
RealMLP	5.3%	Mix Bernoulli Naive Bayes (Mix BernoulliNB)	28.7%
Logistic Regression (LR)	5.3%	Linear Discriminant Analysis (LDA)	37.7%
Linear Support Vector Machine (SVM_Linear)	5.7%	Quadratic Discriminant Analysis (QDA)	45.7%
Polynomial Support Vector Machine (SVM_Poly)	7.9%	Sigmoid Support Vector Machine (SVM_Sigmoid)	58.1%
Multinomial Naive Bayes (MultinomialNB)	7.9%	Dummy Classifier (Dummy)	60%
Nearest Centroid (NC)	7.9%	TabNet	74.3%

Feature Scaling Rarely Mitigate Ill-Behavior

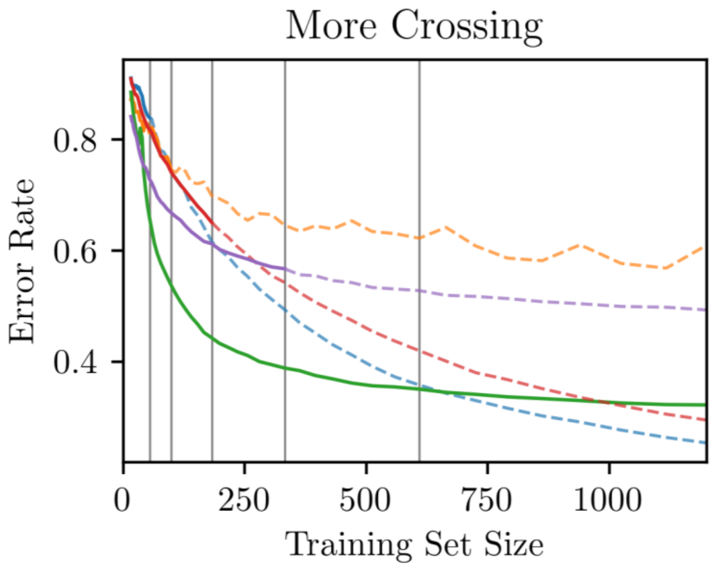
Shapes	LCDB 1.1		
	no feature scaling	min-max scaling	standardization scaling
Non-Monotone ($\neg$ M)	9.6%	11.1%	9.5%
Non-Convex ($\neg$ C)	12.3%	10.0%	8.8%
Ill-behaved ($\neg$ M $\cup$ $\neg$ C)	15.4%	14.3%	11.8%

Downstream Tasks

- **Learning curves modeling** is significantly more difficult for non-monotone and non-convex curves.

- Crossing learning curves make **model selection** (Successive Halving) significantly more challenging.

- …



Conclusion

- **Dataset:** the largest, reproducible, high-resolution collection of learning curves with three preprocessing variants.
- **Observation:** 15% of curves display ill-behavior, depends strongly on which learner is used. Ill-behavior found is robust to feature scaling.
- **Benchmark:** a challenging benchmark for studying ill-behavior and evaluating downstream tasks that rely on learning curves.

LCDB 1.1: A Database Illustrating Learning Curves Are More Ill-Behaved Than Previously Thought

Paper

Code

Dataset